

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2001-188668
(P2001-188668A)

(43) 公開日 平成13年7月10日 (2001.7.10)

(51) Int.Cl.	識別記号	F I	テーマコード(参考)
G 0 6 F 3/16	3 2 0	G 0 6 F 3/16	3 2 0 B 5 B 0 9 1
9/44	5 7 0	9/44	5 7 0 A 5 D 0 1 5
17/28		15/38	X 5 D 0 4 5
G 1 0 L 13/00		G 1 0 L 3/00	R 9 A 0 0 1
15/00			5 5 1 P

審査請求 未請求 請求項の数 6 O L (全 13 頁) 最終頁に続く

(21) 出願番号 特願平11-375772

(22) 出願日 平成11年12月28日 (1999. 12. 28)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川 6 丁目 7 番 35 号

(72) 発明者 浅野 康治

東京都品川区北品川 6 丁目 7 番 35 号 ソニー株式会社内

(72) 発明者 青柳 誠一

東京都品川区北品川 6 丁目 7 番 35 号 ソニー株式会社内

(74) 代理人 100082131

弁理士 稲本 義雄

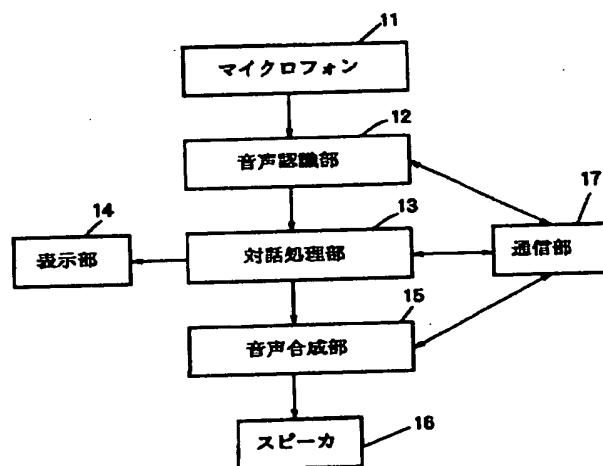
最終頁に続く

(54) 【発明の名称】 会話処理装置および方法、情報処理装置、会話処理システム、並びに記録媒体

(57) 【要約】

【課題】 ネットワークに接続された機器を音声により制御する。

【解決手段】 ユーザは、ネットワークに接続されている機器を操作する場合、音声対話装置 1 のマイクロフォン 11 に向かって、所望の操作について発話する。入力されたユーザの発話は、音声認識部 12 により認識される。対話処理部 13 は、認識結果に従った操作が行えるような情報を生成する。生成された情報は、通信部 17 に出力され、さらにネットワークを介して、操作対象とされた機器に送信される。機器は、ネットワークを介して送信されてきた情報に従った処理を実行する。



音声対話装置 1

【特許請求の範囲】

【請求項1】 少なくとも1以上の機器が接続されるネットワークに接続される会話処理装置において、ユーザの発話を認識する認識手段と、前記ネットワークに接続されている前記機器を操作するためのシナリオを記憶する記憶手段と、前記認識手段による認識結果に対応するシナリオを前記記憶手段から読み出す読み出し手段と、前記読み出し手段により読み出された前記シナリオに従って、前記ユーザと会話し、その会話のなかから、前記機器を操作するための操作情報を取得する取得手段と、前記取得手段により取得された前記操作情報を前記ネットワークを介して前記機器に送信する送信手段とを含むことを特徴とする会話処理装置。

【請求項2】 前記ネットワークに接続されている機器の接続状態を管理する管理手段と、前記管理手段により、前記ネットワークに接続されている機器の接続状態に変更があったと判断された場合、前記ネットワークに接続されている機器に対して、記憶しているシナリオを出力するように要求する要求手段とをさらに含み、前記記憶手段は、前記要求手段により要求した結果、前記ネットワークに接続されている機器から出力された前記シナリオを記憶することを特徴とする請求項1に記載の会話処理装置。

【請求項3】 少なくとも1以上の機器が接続されるネットワークに接続される会話処理装置の会話処理方法において、ユーザの発話を認識する認識ステップと、前記認識ステップの処理による認識結果に対応するシナリオに従って、前記ユーザと会話し、その会話のなかから、前記機器を操作するための操作情報を取得する取得ステップと、前記取得ステップの処理で取得された前記操作情報の前記ネットワークを介する前記機器への送信を制御する送信制御ステップとを含むことを特徴とする会話処理方法。

【請求項4】 少なくとも1以上の機器が接続されるネットワークに接続される会話処理装置の会話処理用のプログラムであって、ユーザの発話を認識する認識ステップと、前記認識ステップの処理による認識結果に対応するシナリオに従って、前記ユーザと会話し、その会話のなかから、前記機器を操作するための操作情報を取得する取得ステップと、前記取得ステップの処理で取得された前記操作情報の前記ネットワークを介する前記機器への送信を制御する送信制御ステップとを含むことを特徴とするコンピュータが読み取り可能なプログラムが記録されている記録媒体。

【請求項5】 少なくとも1以上の機器が接続されるネットワークに接続される情報処理装置において、音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報を記憶する記憶手段と、

前記ネットワークに接続されている他の装置から、前記記憶手段に記憶されている情報の出力要求を受けた場合、前記ネットワークを介して、前記他の装置に対して前記情報を出力する出力手段とを含むことを特徴とする情報処理装置。

【請求項6】 会話処理装置と情報処理装置とで構成される会話処理システムにおいて、

前記会話処理装置は、

ユーザの発話を認識する認識手段と、

前記ネットワークに接続されている前記情報処理装置を操作するためのシナリオを記憶する第1の記憶手段と、前記認識手段による認識結果に対応するシナリオを前記第1の記憶手段から読み出す読み出し手段と、

前記第1の読み出し手段により読み出された前記シナリオに従って、前記ユーザと会話し、その会話のなかから、前記情報処理装置を操作するための操作情報を取得する取得手段と、

前記取得手段により取得された前記操作情報を前記ネットワークを介して前記情報処理装置に送信する送信手段とを含み、

前記情報処理装置は、

前記会話処理装置を介して音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報を記憶する第2の記憶手段と、

前記会話処理装置から、前記第2の記憶手段に記憶されている情報の出力要求を受けた場合、前記ネットワークを介して、前記会話処理装置に対して前記情報を出力する出力手段と、

前記送信手段により送信された前記操作情報を受信し、それに従った操作を実行する実行手段とを含むことを特徴とする会話処理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は会話処理装置および方法、情報処理装置、会話処理システム、並びに記録媒体に関し、特に、ネットワークに接続した機器を音声により操作するようなシステムに用いて好適な会話処理装置および方法、情報処理装置、会話処理システム、並びに記録媒体に関する。

【0002】

【従来の技術】 近年、ネットワーク化が普及しつつある。家庭内においても、さまざま機器が、例えば、IEEE 1394などのデジタルケーブルで相互に接続されることにより、ネットワークが構築され、データの授受を行えようになっている。また、従来、手動にて行われていた機

器の操作が、音声により行える機器も普及しつつある。

【0003】

【発明が解決しようとする課題】家庭内において、ネットワークが構築、機器同士が接続されている場合、それらの機器で、データを相互に授受するときは、データを送信する側または受信する側、どちらかにおいて操作を行うか、パーソナルコンピュータなどにより、ネットワークを管理し、そのパーソナルコンピュータにより操作を行う。どちらの操作方法によっても、操作は、手で行われることが多かった。その為、操作に慣れないユーザは、手間取るという課題があった。

【0004】また、音声により操作できる機器も、その機器の近辺にいるときしか、その装置を操作できないといった課題があった。換言すれば、ネットワークで他の装置と接続されていても、目の前にある装置を用いて、音声により他の装置を操作するということは困難であった。

【0005】本発明はこのような状況に鑑みてなされたものであり、ネットワークに接続されている機器を、一括して音声により操作できるようにすることにより、ユーザに使い勝手の良いネットワーク環境を提供することを目的とする。

【0006】

【課題を解決するための手段】請求項1に記載の会話処理装置は、ユーザの発話を認識する認識手段と、ネットワークに接続されている機器を操作するためのシナリオを記憶する記憶手段と、認識手段による認識結果に対応するシナリオを記憶手段から読み出す読み出し手段と、読み出し手段により読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、機器を操作するための操作情報を取得する取得手段と、取得手段により取得された操作情報をネットワークを介して機器に送信する送信手段とを含むことを特徴とする。

【0007】ネットワークに接続されている機器の接続状態を管理する管理手段と、管理手段により、ネットワークに接続されている機器の接続状態に変更があったと判断された場合、ネットワークに接続されている機器に対して、記憶しているシナリオを出力するように要求する要求手段とをさらに含み、記憶手段は、要求手段により要求した結果、ネットワークに接続されている機器から出力されたシナリオを記憶するようにすることができ

る。

【0008】請求項3に記載の会話処理方法は、ユーザの発話を認識する認識ステップと、認識ステップの処理による認識結果に対応するシナリオに従って、ユーザと会話し、その会話のなかから、機器を操作するための操作情報を取得する取得ステップと、取得ステップの処理で取得された操作情報のネットワークを介する機器への送信を制御する送信制御ステップとを含むことを特徴とする。

【0009】請求項4に記載の記録媒体のプログラムは、ユーザの発話を認識する認識ステップと、認識ステップの処理による認識結果に対応するシナリオに従って、ユーザと会話し、その会話のなかから、機器を操作するための操作情報を取得する取得ステップと、取得ステップの処理で取得された操作情報のネットワークを介する機器への送信を制御する送信制御ステップとを含むことを特徴とする。

【0010】請求項5に記載の情報処理装置は、音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報を記憶する記憶手段と、ネットワークに接続されている他の装置から、記憶手段に記憶されている情報の出力要求を受けた場合、ネットワークを介して、他の装置に対して情報を出力する出力手段とを含むことを特徴とする。

【0011】請求項6に記載の会話処理システムの会話処理装置は、ユーザの発話を認識する認識手段と、ネットワークに接続されている情報処理装置を操作するためのシナリオを記憶する第1の記憶手段と、認識手段による認識結果に対応するシナリオを第1の記憶手段から読み出す読み出し手段と、第1の読み出し手段により読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、情報処理装置を操作するための操作情報を取得する取得手段と、取得手段により取得された操作情報をネットワークを介して情報処理装置に送信する送信手段とを含み、情報処理装置は、会話処理装置を介して音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報を記憶する第2の記憶手段と、会話処理装置から、第2の記憶手段に記憶されている情報の出力要求を受けた場合、ネットワークを介して、会話処理装置に対して情報を出力する出力手段と、送信手段により送信された操作情報を受信し、それに従った操作を実行する実行手段とを含むことを特徴とする。

【0012】請求項1に記載の会話処理装置、請求項3に記載の会話処理方法、および請求項4に記載の記録媒体においては、ユーザの発話が認識され、ネットワークに接続されている機器を操作するためのシナリオが記憶され、認識結果に対応するシナリオが読み出され、その読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、機器を操作するための操作情報が取得され、その取得された操作情報がネットワークを介して機器に送信される。

【0013】請求項5に記載の情報処理装置においては、音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報が記憶され、ネットワークに接続されている他の装置から、記憶されている情報の出力要求を受けた場合、ネットワークを介して、他の装置に対して情報が出力される。

【0014】請求項6に記載の会話処理システムの会話

処理装置においては、ユーザの発話が認識され、ネットワークに接続されている情報処理装置を操作するためのシナリオが記憶され、認識結果に対応するシナリオが読み出され、読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、情報処理装置を操作するための操作情報が取得され、その取得された操作情報がネットワークを介して情報処理装置に送信され、情報処理装置においては、会話処理装置を介して音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報が記憶され、会話処理装置から、記憶されている情報の出力要求を受けた場合、ネットワークを介して、会話処理装置に対して情報が出力され、会話処理装置から送信された操作情報が受信され、それに従った操作が実行される。

【0015】

【発明の実施の形態】以下に、図を参照しながら、本発明の実施の形態について説明する。図1は、本発明を適用した一実施の形態のネットワークの構成を示す図である。

【0016】図1に示した実施の形態では、3つの部屋A乃至Cがあり、部屋Aには、音声対話装置1-1、テレビジョン受像機2、およびVTR（ビデオテープレコーダ）3がある。部屋Bには、音声対話装置1-2とエアーコンディショニング4があり、部屋Cには、パーソナルコンピュータ5がある。これらの機器は、ネットワーク6により、相互にデータの授受が行えるように接続されている。

【0017】ユーザは、音声対話装置1-1または音声対話装置1-2を用いて、音声により、ネットワーク6を介して接続されている機器の操作を行うことができる。なお、ネットワーク6は、有線により構成されても、無線により構成されても、どちらでも良い。また、有線と無線が混在するネットワークでも良い。

【0018】図2は、音声対話装置1-1の構成を示す図である。なお、音声対話装置1-1と音声対話装置1-2は、基本的に同様の構成をしているので、以下においては、音声対話装置1-1と音声対話装置1-2を個々に区別する必要がない場合には、単に、音声対話装置1と記述する。音声対話装置1のマイクロフォン11には、ユーザが発話した音声信号が入力される。マイクロフォン11に入力された音声信号は、音声認識部12により、デジタル信号に変換され、音声認識される。

【0019】音声認識部12により音声認識された認識結果は、テキストデータとして、対話処理部13に出力される。対話処理部13は、入力されたテキストデータを解析し、他の機器を制御するための制御データを生成すると共に、後段の音声合成部15により、音声として出力する応答文のデータを生成する。応答文のデータは、テキストデータであり、必要に応じて、表示部14にも供給され、そのテキストデータに基づくテキストが

表示される。

【0020】音声合成部15は、対話処理部13から供給されたテキストデータを基に、アナログ信号の音声信号を生成し、スピーカ16により発話させる。通信部17は、ネットワーク6を介して他の機器とデータの授受が行えるようになっている。また、通信部17に、電話回線などを介して、外部の機器と接続できるようにしても良い。

【0021】図3は、音声認識部12の詳細な構成を示す図である。ユーザの発話は、マイクロフォン11を介して音声信号として、音声認識部12のAD(Analog Digital)変換部21に入力される。AD変換部21では、アナログ信号である音声信号がサンプリング、量子化され、デジタル信号である音声データに変換される。この音声データは、特徴抽出部22に供給される。

【0022】特徴抽出部22は、AD変換部21からの音声データについて、適当なフレームごとに、例えば、スペクトルや、線形予測係数、ケプストラム係数、線スペクトル対等の特徴パラメータを抽出し、特徴量バッファ23およびマッチング部24に供給する。特徴量バッファ23では、特徴抽出部22からの特徴パラメータが一時記憶される。

【0023】マッチング部24は、特徴抽出部22からの特徴パラメータ、または特徴量バッファ23に記憶された特徴パラメータに基づき、音響モデルデータベース25、辞書データベース26、および文法データベース27を必要に応じて参照しながら、入力された音声信号を認識する。

【0024】即ち、音響モデルデータベース25は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここで、音響モデルとしては、例えば、HMM(Hidden Markov Model)などを用いることができる。辞書データベース26は、認識対象の各単語について、その発音に関する情報が記述された単語辞書を記憶している。文法データベース27は、辞書データベース26の単語辞書に登録されている各単語が、どのように連鎖する(つながる)かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法(CFG)や、統計的な単語連鎖確率(N-gram)などに基づく規則を用いることができる。

【0025】マッチング部24は、辞書データベース26の単語辞書を参照することにより、音響モデルデータベース35に記憶されている音響モデルを接続することによって、単語の音響モデル(単語モデル)を構成する。さらに、マッチング部24は、幾つかの単語モデルを、文法データベース27に記憶された文法規則を参照することによって接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、例えば、HMM法等によって、入力された音声信号を認識する。そして、マッチン

グ部24による音声認識結果は、例えば、テキスト等で出力される。

【0026】なお、マッチング部24は、対話処理部13から、そこで得られる情報を受信することができるようになっており、その会話管理情報に基づいて、精度の高い音声認識を行うことができるようになっている。

【0027】図4は、対話処理部13の詳細な構成を示す図である。音声認識部12から出力された認識結果（テキストデータ）は、対話処理部13の言語処理部31に10 入力される。言語処理部31は、辞書データベース32と解析用文法データベース33に記憶されているデータを基に、入力された音声認識結果を、形態素解析、構文解析などの解析を行うことにより、単語の情報や構文の情報などの言語情報を抽出する。また、辞書に記述された内容を基に、入力の音声発話の意味、意図なども抽出する。

【0028】すなわち、辞書データベース32には、単語の表記や解析用文法を適用するために必要な品詞情報などの情報、単語の個別の意味情報などを記憶されており、解析用文法データベース33には、辞書データベ15 ス32に記憶されている各単語の情報を基に、単語連鎖に関する制約を記述したデータが記憶されている。これらのデータを用いて言語処理部31は、入力された音声認識結果のテキストデータを解析する。

【0029】解析用文法データベース33に記憶されているデータは、正規文法、文脈自由文法、統計的な単語連鎖確立、意味的な解析までを含める場合はHPSGなどの意味論を含んだ言語理論などを用いる、テキスト解析に必要なデータである。

【0030】また、言語処理部31は、対話状況に応じて動的に変換する対話履歴メモリ35に記憶されているユーザとの対話履歴を参照することにより、より精度の高い解析を行えるようになっている。

【0031】言語処理部31により抽出された情報に基づき、対話管理部34は、対話履歴メモリ35に管理されている情報を適宜更新する。また、対話管理部34は、通信部17から送信されてきたシナリオデータ（詳細は後述する）を、シナリオメモリ36に出力し、記憶させる一方で、言語処理部31から供給されたデータ（ユーザが音声により指示したこと）に基づき、シナリオメモリ36に記憶されているシナリオに基づき、他の機器を操作するための操作信号を、信号部17を介して出力する。知識データベース40は、ネットワーク6に接続されている機器を制御するために必要な知識が蓄積されている。

【0032】応答文生成部37は、辞書データベース38と応答文生成規則データベース39に記憶されているデータを参照して適切な応答文（テキストデータ）を生成する。

【0033】辞書データベース38は、応答文を生成す

るのに必要な単語情報が記憶されている。辞書データベース32と辞書データベース38は、同じ情報を記憶するようにしても良い。従って、共用することも可能である。

【0034】会話生成規則データベース39は、どのような応答文を、どのように生成するのかに関する規則が記憶されている。例えば、仮に、話題が意味的なフレーム構造などで管理されているなら、そこから自然言語文を生成する規則などが書かれている。この意味構造から自然言語文を生成する方法は、言語処理部31により行われる処理の逆の処理として行うことにより、実現することが可能である。

【0035】このようにして、応答文生成部37により生成された、テキストデータとしての応答文は、音声合成部15に出力される。

【0036】図5は、音声合成部15の構成例を示している。テキスト解析部51には、対話処理部13が出力するテキストが、音声合成処理の対象として入力されるようになっており、テキスト解析部51は、辞書データベース52や解析用文法データベース53を参照しながら、そのテキストを解析する。

【0037】即ち、辞書データベース52には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、解析用文法データベース53には、辞書データベース52の単語辞書に記述された単語について、単語連鎖に関する制約等の解析用文法規則が記憶されている。そして、テキスト解析部51は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部54で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報、その他の韻律情報や、各単語の発音等の音韻情報などがある。

【0038】テキスト解析部51で得られた情報は、規則合成部54に供給され、規則合成部54では、音素片データベース55を用いて、テキスト解析部51に15 入力されたテキストに対応する合成音の音声データ（デジタルデータ）が生成される。

【0039】即ち、音素片データベース55には、例えば、CV(Consonant, Vowel)や、VCV、CVC等の形で音素片データが記憶されており、規則合成部54は、テキスト解析部51からの情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト解析部51に15 入力されたテキストに対応する合成音の音声データを生成する。

【0040】この音声データは、DA変換部56に供給され、そこで、アナログ信号としての音声信号に変換さ

れる。この音声信号は、スピーカ16に供給され、これにより、テキスト解析部51に入力されたテキストに対応する合成音出力される。

【0041】ここで、対話処理部13のシナリオメモリ36に記憶されているシナリオについて、図6を参照して説明する。シナリオメモリ36には、各機器毎（各機器の操作毎）に、用意されたシナリオが記憶されている。各シナリオは、機器が所定の動作を行うために必要な情報が取得されるように、構成されている。即ち、図6に示した例では、VTR3に録画予約をさせたい場合、“録画日”、“録画開始時間”、“録画終了時間”、“および”チャンネル”が、情報として必要であり、それらの情報が得られるような構成となっている。

【0042】対話制御部34は、VTR3による番組の録画予約が指示された際、シナリオメモリ36内に記憶されている図6に示したようなシナリオを読み出し、各スロットが埋められる方向で、ユーザと対話を進める。

【0043】シナリオメモリ36に記憶されるシナリオは、ネットワーク6に接続されている機器自身が保持し、所定のタイミングにおいて、音声対話装置1のシナリオメモリ36に出力され、記憶される。このシナリオが出力されるタイミングについては、後述する。このように、機器自身が自己のシナリオを保持しているため、各機器は、内部にシナリオを記憶するためのメモリを備えている。

【0044】図7は、ネットワーク6に接続されている機器、図1に示した実施の形態では、テレビジョン受像機2、VTR3、エアーコンディショニング4、パーソナルコンピュータ5などの内部構成を示す図である。機器制御部61は、通信制御部62によりネットワーク6を介して音声対話装置1から入力された制御データに基づき、固有機能処理部63による処理を制御する。固有制御処理部63は、機器の持つ固有の機能などを実現する部分であり、例えば、VTR3の場合、録画や再生など処理を行う機能であり、エアーコンディショニング4の場合、室温調整機能などである。

【0045】情報記憶部64は、機器固有のシナリオや、その機器を音声により操作する際に必要な単語などの情報も記憶されている。

【0046】図8のフローチャートを参照して、音声対話装置1の動作のうち、ネットワーク6に接続されている機器の状態管理について説明する。音声対話装置1は、ステップS1において、ネットワークの接続状態のチェックを行う。ステップS2において、状態に変更があったか否かが判断される。

【0047】ステップS1とステップS2の処理は、前回チェックを行った際に接続されていた機器の個数（個数 $t-1$ とする）を記憶しておき、現在チェックを行った結果、接続されていると判断された機器の個数（個数 t とする）とを比較し、個数 $t-1 > 個数t$ の場合、ネ

ットワークから機器が外されたと判断し、個数 $t-1 < 個数t$ の場合、ネットワークに新たな機器が追加されたと判断し、個数 $t-1 = 個数t$ の場合、ネットワークに接続されている機器の状態に変化はないと判断することにより行われる。

【0048】ステップS2において、ネットワークの機器の接続状態に変化がなかったと判断された場合、ステップS1に戻り、それ以降の処理が繰り返され、ネットワークの機器の接続状態に変化があったと判断された場合、ステップS3に進む。音声対話装置1は、ステップS3において、その時点で、ネットワーク6に接続されている全ての機器に対して、情報記憶部64に記憶されているシナリオおよび情報を転送するように要求する。

【0049】ステップS4において、音声対話装置1内の情報が更新される。まず、音声認識部12の辞書データベース26と文法データベース27（図3）に記憶されている情報が更新される。辞書データベース26に記憶されている情報は、ネットワーク6に接続されている全ての機器から、それぞれの機器を音声で制御する際に必要とされる単語の情報であり、それらの全ての単語を含むような辞書に更新される。

【0050】解析文法データベース27に関しては、例えば、文脈自動文法形式で記述された文法に基づき、辞書データベース26に含まれる単語を用いて発生される発話を受け入れるように更新される。

【0051】対話処理部13のシナリオメモリ36に関しては、ネットワークに接続された全ての機器から、それぞれの機器を音声で制御するためのシナリオ情報を取得し、それらの全てのシナリオを含むように更新される。知識データベース40に関しては、ネットワークに接続された全ての機器から、それぞれの機器を音声で制御するための知識を取得し、それらの全ての知識を含むように更新される。

【0052】音声合成部15の辞書データベース52に関しては、ネットワーク6に接続されている全ての機器から、それぞれの機器を音声で制御する際の応答文の生成に必要な単語の情報を取得し、それらの全ての単語の情報を含むように更新される。解析用文法データベース53に関しては、例えば、文脈自動文法形式で記述された文法に基づき、辞書データベース52に含まれる単語を用いて生成される応答文を受け付けるように更新する。

【0053】図8に示したフローチャートの処理は、常に行われるようにしても良いし、所定の周期で行われようようにしても良い。所定の周期としては、分単位でも良いし、日単位でもよい。また、図8に示したフローチャートにおいては、音声対話装置1が、ネットワーク6に接続されている機器の状態を管理するようにしたが、新たにネットワーク6に接続された機器が、接続された時点で、音声対話装置1に対して、情報記憶部64に記憶し

10

20

30

40

50

ている情報を出力するようにしても良い。

【0054】このように、新たに接続された機器自身が、音声対話装置1に対して、自己の存在を知らせることにより、音声対話装置1は、ネットワーク6に接続された機器の情報を取得することができる。このようにすれば、ネットワーク6に接続されている機器の状態に変更が生じて、全ての機器から情報を収集する必要はなく、新たに接続された機器からのみ情報を取得すれば良いことになる。

【0055】ただし、ネットワーク6に接続されていた機器がネットワーク6から外された場合は、音声対話装置1により検出することはできないので、そのような状態に対応するために、所定の周期でネットワーク6の接続状態を確認する、または、ユーザがネットワーク6から外された機器に対して指示を出した時に、その機器から応答がないことにより、外されたことを認識し、その時点で、その機器に対する情報を削除するなどの処理を行うようにする必要がある。

【0056】次に、図9のフローチャートを参照して、音声対話装置1の動作のうち、ユーザの音声による指示により、ネットワーク6に接続されている機器を制御する場合の動作について説明する。

【0057】ステップS11において、音声対話装置1は、マイクロフォン11（図2）に、ユーザの音声が入力されたか否かを判断する。ステップS11において、音声入力があったと判断された場合、ステップS12に進み、入力された音声に対して、音声認識部12により、音声認識が行われる。音声認識部12により出力された認識結果は、対話処理部13（図4）に入力される。

【0058】ステップS13において、対話処理部13の言語処理部31は、入力された認識結果に対して、言語処理を施し、対話管理部34に処理結果を出力する。ステップS14において、対話管理部34は、対話処理を行う。このようにして、ユーザが発話された言葉は、音声対話装置1により認識され、処理される。そして、ステップS15において、対話管理部34は、処理結果を用いて、シナリオが埋まったか否かを判断し、シナリオは埋まっていないと判断された場合、ステップS16に進み、応答文生成部37により応答文が生成される。

【0059】ここで、具体的な例を挙げ、ステップS11乃至S16の処理について説明する。ユーザが時刻t₁において、“ビデオの録画予約”と音声対話装置1に向かって発話すると、その音声は、マイクロフォン11により音声対話装置1内に取り込まれる。従って、ステップS11において、音声入力ありと判断され、ステップS12に進む。ステップS12において、音声認識部12により、ユーザが、なんと発話したのかが、認識される。その認識結果は、テキストデータとして、対話処理部13の言語処理部31に出力される。

【0060】ステップS13およびS14において、言語処理部31と対話管理部34において、言語処理と対話処理が行われる。その結果、ユーザがVTR3による録画予約を指示したと判断される。そして、対話管理部34は、その判断に基づき、シナリオメモリ36に記憶されている複数のシナリオのうち、図6に示したような“VTR録画予約”というシナリオを読み出す。

【0061】対話管理部34は、読み出したシナリオに従い処理を開始する。すなわち、この場合、ステップS15において、まだ、読み出したシナリオの4つのスロットは、1つも埋まっていないと判断し、ステップS16に進み、埋まっていないスロットを埋める方向で、ユーザと対話されるように、応答文生成部37により応答文を生成させる。

【0062】この際、対話管理部34は、対話履歴メモリ35や知識データベース40も参照し、この場合、例えば、“録画予約ですね。録画日はいつですか？”といった応答文が応答文生成部37により生成されるための情報を出力する。そして、応答文生成部37により応答文となるテキストデータが生成され、音声合成部15により音声合成され、スピーカ16によりユーザに対して生成された応答文が発話される。

【0063】このような処理が繰り返されることにより、図6に示した“VTR録画予約”というシナリオの各スロットが埋められると、ステップS15において、シナリオは埋まったと判断され、ステップS17に進む。ステップS17において、対話処理結果を対話履歴メモリ35と知識データベース40に反映する。このように、対話処理結果を対話履歴メモリ35と知識データベース40に反映することにより、例えば、エアーコンディショニング4に対する過去の操作履歴を用いて、その履歴に設定室温が保持されていれば、ユーザによりエアーコンディショニング4のスイッチをオンにするという指示が出された場合、その保持されている設定室温をデフォルトして用いるといったことが可能となる。

【0064】ステップS18において、通信部17を介して、ネットワーク6に接続されている機器を制御する。例えば、VTR3に対する録画予約の指示が、ユーザにより行われている場合、その予約の設定（シナリオのスロットに書かれた情報が出力）が行われる。そして、ステップS19において、応答文が生成され、ステップS20において、スピーカ16により生成された応答文が発話される。

【0065】例えば、VTR3に対して、録画予約の設定を行うように指示したが、その指示された時間帯には、既に別の予約が入っていると判断された場合、そのことをユーザに知らせる応答文が生成される。別の予約が入っているか否かの判断は、VTR3が、録画予約が指示された時に、自己に保持している録画予約のリストを参照し、重なる時間帯があったら、音声対話装置1に、その

13

ことを知らせるようにしても良いし、音声対話装置1自体が、対話履歴メモリ35に記憶されている過去の対話の履歴を参照し、重なる時間帯に録画予約が指示された履歴が残っているか否かを調査することにより行われるようにしても良い。

【0066】VTR3において、別の予約が入っているか否かを判断する場合は、ステップS18において、まず、VTR3に対して、録画予約が指示され、ステップS19において、指示を受けたVTR3からの返答に基づき、応答文が作成される。また、VTR3は、録画指示を受けた際、録画用のテープが装着されていないような場合、そのことを音声対話装置1に知らせ、音声対話装置1は、そのことをユーザに知らせるための応答文を生成する。

【0067】音声対話装置1において、別の予約が入っているか否かを判断する場合は、ステップS18の処理とステップS19の処理を入れ替えるようにしても良い。即ち、音声対話装置1により別の予約が入っていないと判断されると、例えば、“予約を行います”という応答文が生成され、その後、実際に、VTR3に対して録画予約の指示が出される。

【0068】なお、ステップS20において、音声として応答文が出力されるとともに、表示部14にテキストとして表示されるようにしても良い。また、例えば、ユーザにより設定された予約日時やチャンネルなどの情報を表示部14に表示するようにしても良い。

【0069】一方、ステップS11において、音声入力はないと判断された場合、ステップS21に進み、ネットワーク6に接続されている機器から割り込みがあるか否かが判断される。機器からの割り込みとしては、例えば、パーソナルコンピュータ5に、電子メールが着信し、そのことをユーザに知らせるための割り込みや、お風呂の湯加減を管理する装置が接続されている時には、その装置から、設定温度になったことを知らせるための割り込みなどがある。

【0070】ステップS21において、機器からの割り込みがあったと判断された場合、ステップS14以下の処理において、その割り込みに対応するための処理が行われ、割り込みがないと判断された場合、ステップS11に戻り、それ以降の処理が繰り返される。

【0071】上述した説明においては、図1に示したように、1つのネットワーク6に、音声対話装置1-1と音声対話装置1-2が接続されている。このように、1つのネットワークに複数の音声対話装置1が接続されている場合、それらの複数の音声対話装置1は、全て同様の構成としても良いし、共有できる部分は共有するようにしても良い。

【0072】例えば、音声対話装置1-1を親機と位置づけ、音声対話装置1-2を子機と位置づけた場合、親機としての音声対話装置1-1は、図2に示したような

14

内部構成にする。そして、子機としての音声対話装置1-2は、マイクロフォン11、スピーカ16、および通信部17（以下、親機としての音声対話装置1-1に備えられているマイクロフォン11、スピーカ16、および通信部17と区別できるように、それぞれダッシュを付けて記述する）とから構成されるようにする。

【0073】このような構成にした場合、音声対話装置1-2のマイクロフォン11'により取り込まれたユーザの音声は、デジタル信号に変換されて、通信部17'により、ネットワーク6を介して音声対話装置1-1に送信される。そして、音声対話装置1-1において、上述したような処理が行われ、その結果としての、応答文を発話するためのデータのみが、ネットワーク6を介して音声対話装置1-2に送信され、そのデータに基づいて、音声対話装置1-2のスピーカ16'により応答文が発話される。

【0074】複数の音声対話装置1を同様な構成とした場合、図9に示したフローチャートにおいて、ステップS17の処理は、ネットワーク7に接続されている全ての音声対話装置1に対する処理となる。即ち、音声対話装置1-1において、図9に示したフローチャートの処理が実行されているとき、ステップS17の処理は、ネットワーク6に接続されている全ての音声対話装置1に対して、対話処理結果を送信する処理となる。そして、それぞれの音声対話装置1において、受信された対話処理結果が、対話履歴メモリ35と知識データベース40内のデータに、反映されるような処理が実行される。

【0075】また、複数の音声対話装置1を同様な構成とした場合、図8に示したフローチャートの処理は、それぞれの音声対話装置1により行われるようにしても良いし、1台が代表して行うようにしても良い。1台が代表して行うようにした場合、ネットワーク6に接続されている機器の状態に変化があり、情報が更新されたとき、そのことを知らせるデータ（更新された情報を含めても良い）を、他の音声対話装置1に送信する。すなわち、図8に示したフローチャートのうち、ステップS4の処理の後に、ステップS5として、他の音声対話装置1に対して、情報が更新されたことを知らせるという処理を設ければ良い。

【0076】上述した実施の形態においては、音声会話装置1は、直接ユーザの音声をマイクロフォン11により入力するとして説明したが、電話回線などを介して送信されてきた音声信号も入力されるようにしても良い。そのような構成にした場合、外部から家庭内のネットワーク6に接続されている機器を操作することが可能となる。

【0077】また、音声会話装置1の対話履歴メモリ35や知識データベース40などに記憶されているデータなどは、ユーザ毎に異なり、そのようなデータを記憶する部分を音声会話装置1に対して着脱自在にし、その部

15

分をユーザは携帯することにより、外出先などにおいても、家庭内と同じ操作（ユーザに適した操作）を行うことが可能となる。

【0078】さらに、上述した実施の形態においては、音声会話装置1は、1つの装置として説明したが、例えば、テレビジョン受像機2やVTR3などの機器に組み込まれるような構成としても良い。

【0079】上述した一連の処理は、ハードウェアにより実行させることもできるが、ソフトウェアにより実行させることもできる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、記録媒体からインストールされる。

【0080】この記録媒体は、図10に示すように、コンピュータとは別に、ユーザにプログラムを提供するために配布される、プログラムが記録されている磁気ディスク111（フロッピディスクを含む）、光ディスク112（CD-ROM（Compact Disk-Read Only Memory）、DVD（Digital Versatile Disk）を含む）、光磁気ディスク113（MD（Mini-Disk）を含む）、若しくは半導体メモリ114などよりなるパッケージメディアにより構成されるだけでなく、コンピュータに予め組み込まれた状態でユーザに提供される、プログラムが記憶されているROM102や記憶部108が含まれるハードディスクなどで構成される。

【0081】なお、本明細書において、媒体により提供されるプログラムを記述するステップは、記載された順序に従って、時系列的に行われる処理は勿論、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。

【0082】また、本明細書において、システムとは、複数の装置により構成される装置全体を表すものである。

【0083】

【発明の効果】以上の如く請求項1に記載の会話処理装置、請求項3に記載の会話処理方法、および請求項4に記載の記録媒体によれば、ユーザの発話を認識し、ネットワークに接続されている機器を操作するためのシナリオを記憶し、認識結果に対応するシナリオを読み出し、その読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、機器を操作するための操作情報を取得し、その取得された操作情報をネットワークを介して機器に送信するようにしたので、音声によりネットワークに接続されている機器を操作することが可能となる。

【0084】また、請求項5に記載の情報処理装置によれば、音声により操作された場合に必要となるシナリオ

16

の情報、単語の情報、および文法の情報を記憶し、ネットワークに接続されている他の装置から、記憶されている情報の出力要求を受けた場合、ネットワークを介して、他の装置に対して情報を出力するようにしたので、音声で、ネットワークにより接続されている他の装置により操作されることが可能となる。

【0085】さらに、請求項6に記載の会話処理システムの会話処理装置によれば、ユーザの発話を認識し、ネットワークに接続されている情報処理装置を操作するためのシナリオを記憶し、認識結果に対応するシナリオを読み出し、読み出されたシナリオに従って、ユーザと会話し、その会話のなかから、情報処理装置を操作するための操作情報を取得し、その取得された操作情報をネットワークを介して情報処理装置に送信し、情報処理装置によれば、会話処理装置を介して音声により操作された場合に必要となるシナリオの情報、単語の情報、および文法の情報を記憶し、会話処理装置から、記憶されている情報の出力要求を受けた場合、ネットワークを介して、会話処理装置に対して情報を出力し、会話処理装置から送信された操作情報を受信し、それに従った操作を実行するようにしたので、音声で、会話処理装置により、ネットワークに接続されている情報処理装置を操作することが可能となる。

【図面の簡単な説明】

【図1】本発明を適用したネットワークシステムの一実施の形態の構成を示す図である。

【図2】図1の音声対話装置1の内部構成を示す図である。

【図3】図2の音声認識部12の内部構成を示す図である。

【図4】図2の対話処理部13の内部構成を示す図である。

【図5】図2の音声合成部15の内部構成を示す図である。

【図6】シナリオメモリ36に記憶されるシナリオについて説明する図である。

【図7】ネットワークに接続される機器の内部構成を示す図である。

【図8】音声対話装置1の動作について説明するフローチャートである。

【図9】音声対話装置1の他の動作について説明するフローチャートである。

【図10】媒体を説明する図である。

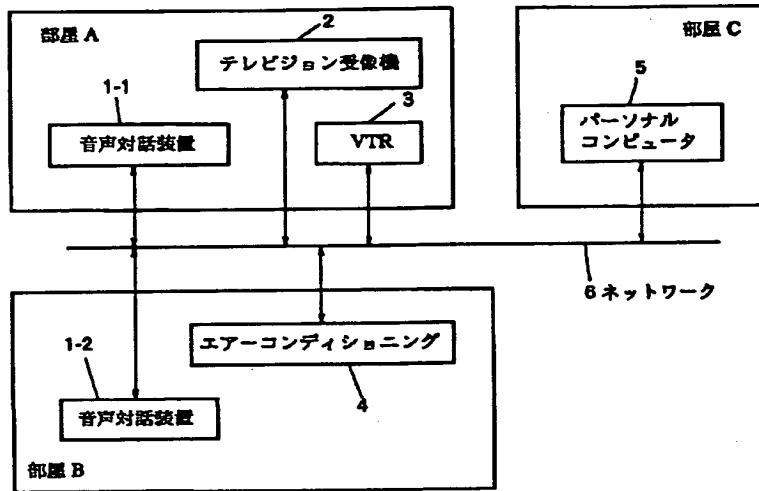
【符号の説明】

1 音声対話装置, 11 マイクロフォン, 12 音声認識部, 13 対話処理部, 14 表示部, 15 音声合成部, 16 スピーカ, 17 通信部, 34 対話管理部, 35 対話履歴メモリ, 36 シナリオメモリ, 40 知識データベース, 61 機器制御部, 62 通信制御部, 63 固有機能処理

17

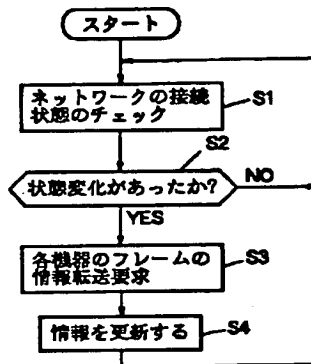
部, 64 情報記憶部

【図1】

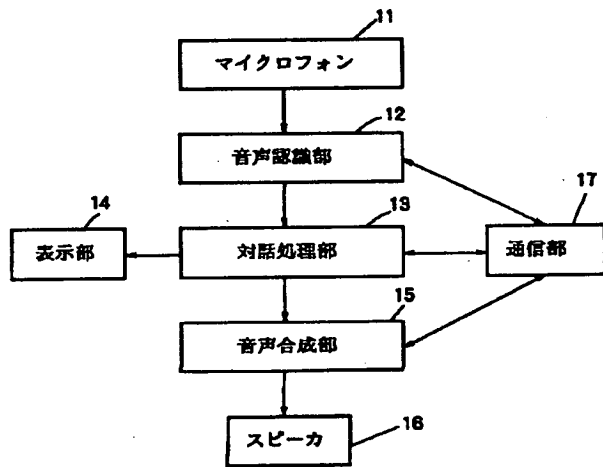


18

【図8】

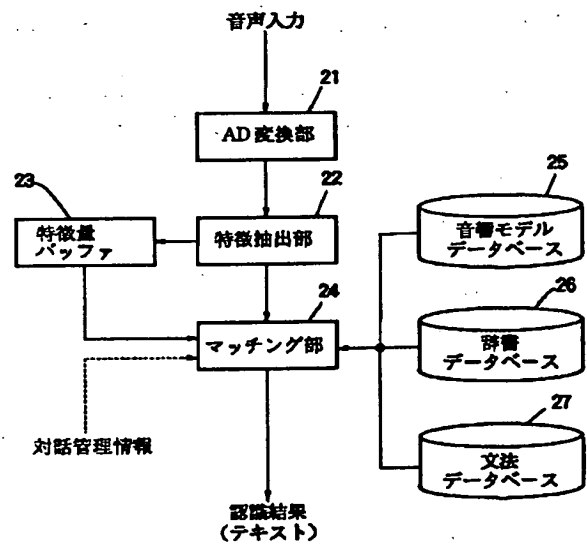


【図2】



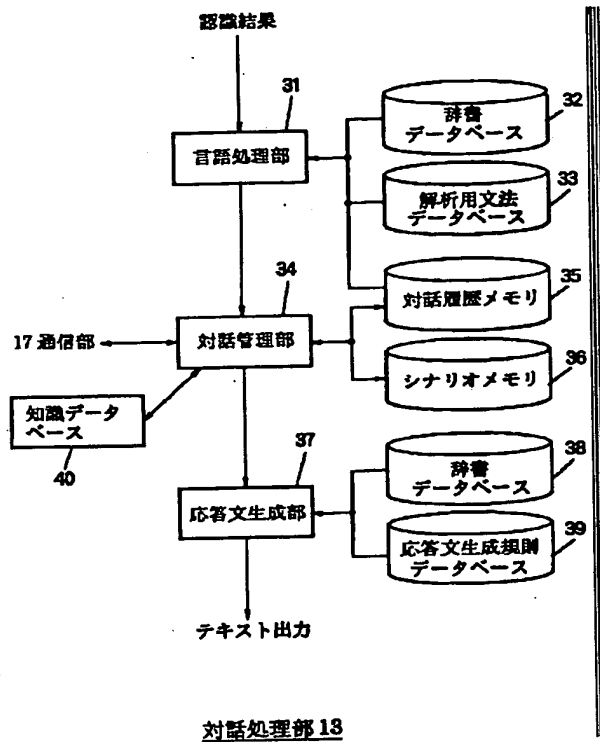
音声対話装置1

【図3】

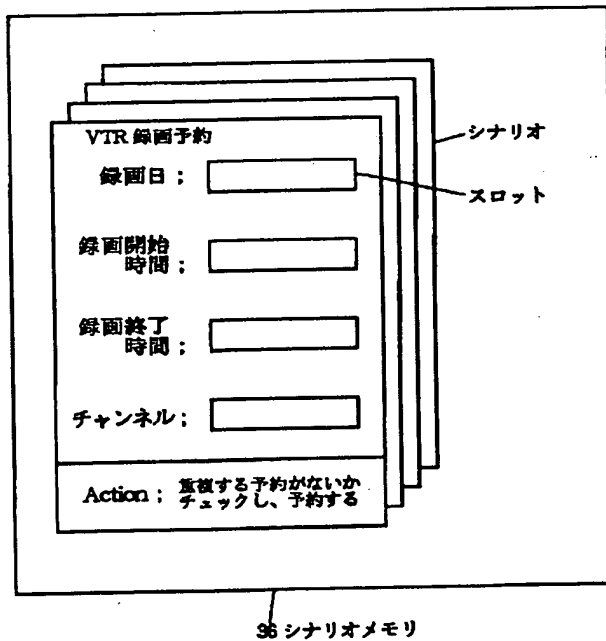


音声認識部12

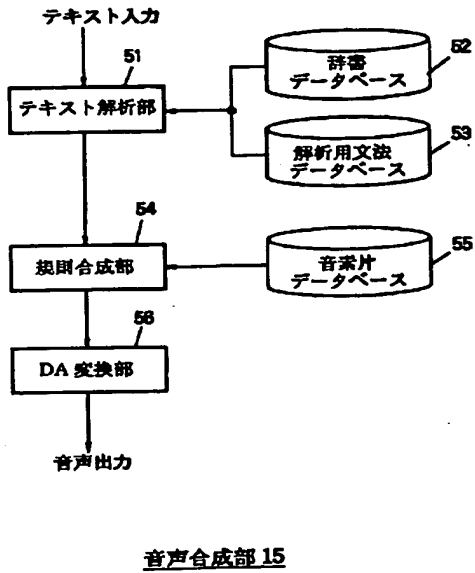
【図4】



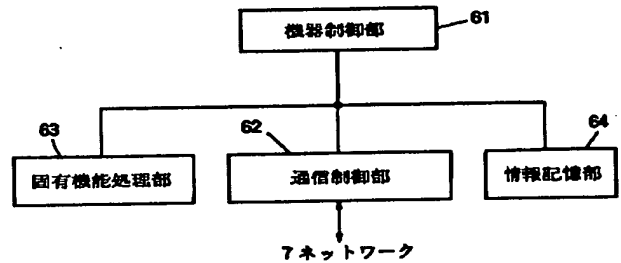
【図6】



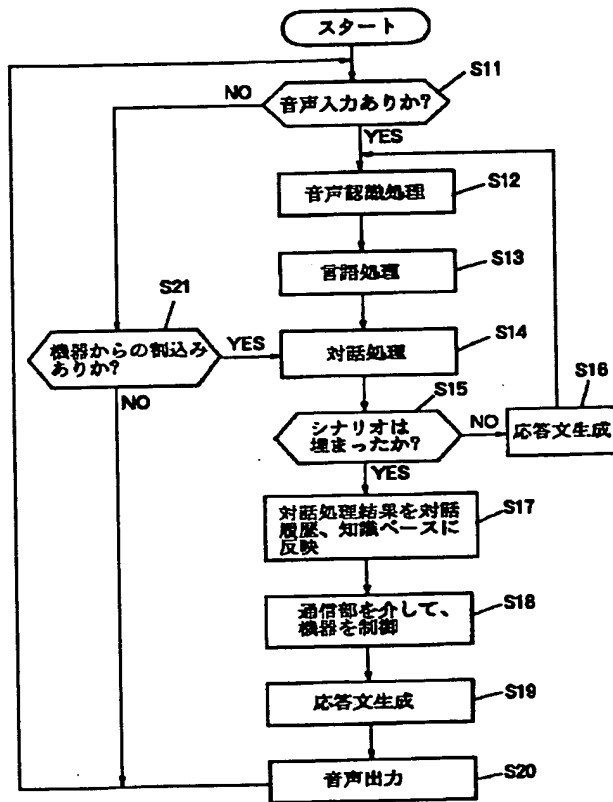
【図5】



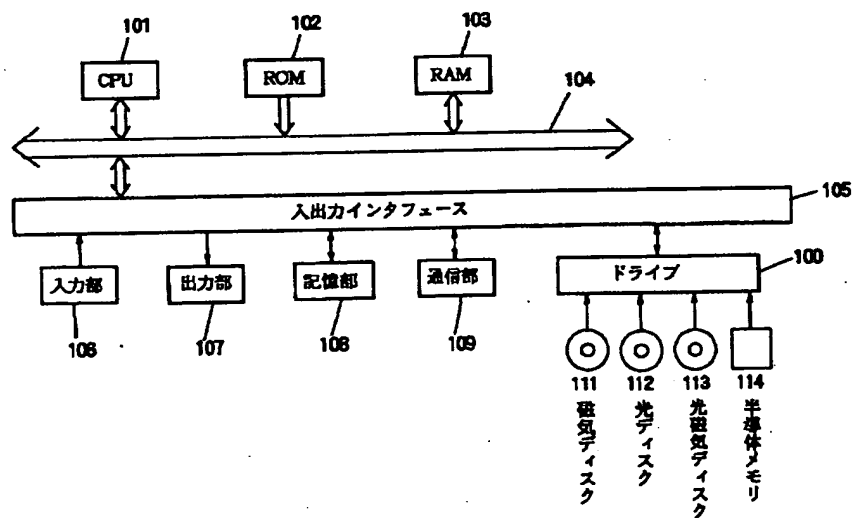
【図7】



【図9】



【図10】



フロントページの続き

(51)Int.Cl. 7 識別記号
G 1 0 L 15/28
15/22

F I
G 1 0 L 3/00

テーマコード(参考)

5 5 1 F
5 7 1 U

(72)発明者 田中 幸
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内
(72)発明者 横野 順
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72)発明者 大江 敏生
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

Fターム(参考) 5B091 AA15 CA02 CA05 CA14 CB12
CB24 CB32 CC01 CC15 CD03
5D015 KK02 KK04 LL01 LL05 LL06
5D045 AB30
9A001 BB06 CC03 CC07 DD11 HH17
HH18 JJ75 KK62